Supplemental Information for

Computational Models for Drug Inhibition of the Human Apical Sodium-dependent Bile Acid Transporter

Xiaowan Zheng[†], Sean Ekins[†,‡,§], Jean-Pierre Raufman[‖], and James E. Polli*[,†]

[†]Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, 20 Penn Street, Baltimore, MD 21201.

[‡] Collaborations in Chemistry, 601 Runnymede Avenue, Jenkintown, PA 19046.

[§] Department of Pharmacology, University of Medicine & Dentistry of New Jersey (UMDNJ)-Robert Wood Johnson Medical School, 675 Hoes Lane, Piscataway, NJ 08854.

[‖]Department of Medicine, Division of Gastroenterology and Hepatology, University of Maryland School of Medicine, 22 South Greene Street, Baltimore, MD 21201.

*Author to whom correspondence should be addressed: Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, HSF2, room 623, Baltimore, MD 21201. Tel: 410-706-8292. Fax: 410-706-5017. E-mail: jpolli@rx.umaryland.edu.

Supplemental Table S1. Training set for the qualitative HipHop model with settings for alignment.

MOLNAME	Activ	Uncert	MaxOmitFeat	Principal
Bendroflumethiazide	92.74	3	0	1
Amlodipine	42.06	3	0	1
Bumetanide	225.2	3	0	0
Dibucaine	34.74	3	0	1
Indomethacin	62.27	3	0	1
Mesoridazine	17.64	3	0	2
Probenecid	385.2	3	0	0
Thioridazine	33.6	3	0	1
Quinine	243.7	3	0	0
Althiazide	500	3	0	0
Triclomethiazide	377.43	3	0	0

Supplemental Table S2. Fifty-eight retrieved compounds from SCUT 2008 database search by using the qualitative HipHop pharmacophore with mesoridazine shape restriction.

Drug	Fit value	Action
Thiothixene	3.758	Antipsychotic
Indomethacin	3.254	inhibits prostaglandin synthesis
Candesartan	3.245	angiotensin II antagonist
Doxorubicin	3.142	intercalates DNA, inhibits DNA topoisomerases I and II
Dipivefrin	3.141	alpha adrenergic agonist
Enalapril	3.049	ACE inhibitor
Aztreonam	2.973	inhibits bacterial cell wall biosynthesis
Bumetanide	2.929	loop diuretic inhibits reabsorption of NA and Cl in the ascending loop of Henle and the distal renal tubule
Thiethylperazine	2.896	antidopaminergic, antiemetic
Mesoridazine	2.871	Antipsychotic
Nimodipine	2.829	calcium channel blocker
Amlodipine	2.82	Calcium channel blocker
Sertaconazole	2.777	Antifungal
Losartan	2.582	angiotensin II antagonist
Lansoprazole	2.566	proton pump inhibitor
Fluvastatin	2.565	HMG-CoA reductase inhibitor
Flurazepam_Metabolite_2_Hydroxyethyl	2.549	.
Nafcillin	2.492	bactericidal, inhibits cell wall synthesis
Rabeprazole Sodium	2.432	proton pump inhibitor
Eprosartan	2.413	angiotensin II receptor antagonist
Daunorubicin	2.338	DNA intercalating agent, inhibits topoisomerase II, generates oxygen free

		radicals
Bortezomib	2.327	Proteasome inhibitor
Thioridazine	2.314	Antipsychotic
Torasemide	2.277	inhibits reabsorption of sodium and chloride in ascending loop of Henle and distil tubule
Aripiprazole	2.26	dopamine and serotonin antagonist
Latanoprost	2.047	Prostaglandin
Mycophenolic Acid	2.016	inhibits immunologically mediated inflammatory response
Glyburide	1.985	stimulates pancreatic insulin release, increases peripheral insulin sensitivity, decreases hepatic glucose output and production, decreases intestinal absorption of glucose
Sulfinpyrazone	1.96	inhibits renal tubular absorption of uric acid
Hydroxyzine	1.936	antihistamine, anxiety
Midazolam_Metabolite_Alpha_Hydroxy	1.921	.
Topiramate	1.901	Anticonvulsant
Alprostadil	1.851	prostaglandin E1 - vasodilator, platelet aggregation inhibitor
Lovastatin	1.733	HMG-CoA reductase inhibitor
Midazolam	1.672	short acting benzodiazepine
Eszopiclone	1.623	nonbenzodiazepine hypnotic
Eszopiclone	1.623	.
Pentamidine	1.541	inhibits DNA, RNA, phospholipid and protein synthesis
Simvastatin	1.503	HMG-CoA reductase inhibitor
Lorazepam	1.403	benzodiazepine, antianxiety
Pravastatin	1.344	HMG-CoA reductase inhibitor
Dinoprostone	1.281	prostaglandin, induces uterine contraction
Celecoxib	1.24	COX-2 inhibitor, NSAID
Bimatoprost	1.226	.

Oxacillin	1.061	bactericidal, inhibits cell wall synthesis
Furosemide	0.959	loop diuretic, inhibits sodium and chloride reabsorption in ascending loop of Henle and distal tubule
Tadalafil	0.932	.
Pioglitazone	0.688	increases insulin sensitivity by PPAR inhibition
Alprazolam_Metabolite_4_Hydroxy	0.646	.
Ezetimibe	0.59	HMG-CoA reductase inhibitor
Ezetimibe	0.586	.
Clorazepate	0.519	Antianxiety
Tioconazole	0.5	topical antifungal
Pramoxime	0.329	topical anesthetic
Ofloxacin	0.314	bactericidal, inhibits DNA gyrase
Tadalafil	0.173	PDE5 inhibitor
Benazepril	0.158	Angiotensin converting enzyme inhibitor
Heroin	0.081	.

Supplemental Table S3. Summary table for the Bayesian model. An ROC plot was generated and the area under the curve (XV ROC AUC) calculated. Best split was identified by the sum of the percent misclassified for category members and for category nonmembers, using the cross-validated score for each sample. Using that split, a contingency table is constructed, containing the number of true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN).

XV ROC AUC	**Best Split**	**TP/FN** **FP/TN**	**# in Category**
0.908	-1.085	15/2 2/19	17

Supplemental Table S4. Bayesian model enrichment results.

This table shows the output name, the percentage of samples that are in that particular category, the number of category members, and the percentage of true members found. Percentages that are less than 100% are in **bold**.

Category %	**1%**	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
44.737%	**5.9%**	**11.8%**	**23.5%**	**52.9%**	**94.1%**	**94.1%**	**94.1%**	**94.1%**	100%

Supplemental Table S5. Percentile results.

This table shows, for each model, the cutoff needed to capture a particular percentage of the good samples. For each cutoff, it shows below the estimated percentages of false positives and true negatives for the non-good samples. This table is designed to identify the cutoff value that best balances the desire to capture as many good samples as possible, while minimizing the number of false positives. The rates shown in this table are estimates derived from the cross-validated data. Cutoff which lead to 10% or greater false positives are displayed in **bold** for ease of identification.

99%	**95%**	**90%**	**70%**	**50%**	**30%**	**10%**	**5%**	**1%**
-13.576	-7.739	-4.570	-1.068	-1.068	13.106	16.608	19.777	25.614
99%/1%	**78%**/22%	**49%**/51%	**18%**/82%	1%/99%	1%/99%	1%/99%	1%/99%	1%/99%

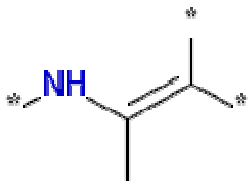
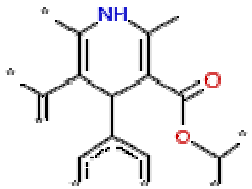
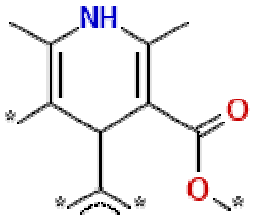
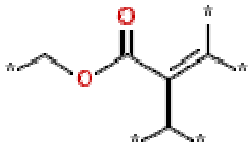
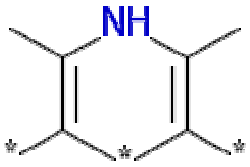
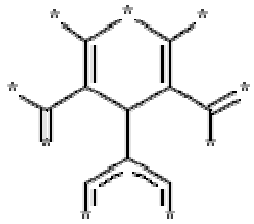
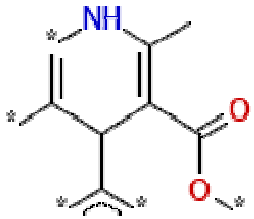
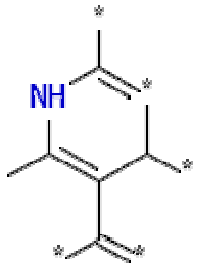
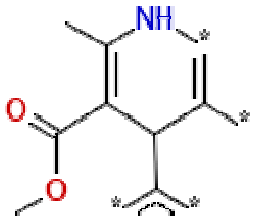
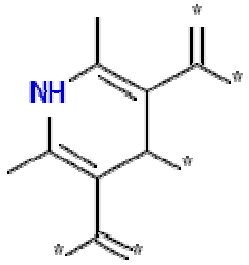
Supplemental Table S6. Category statistics summary.

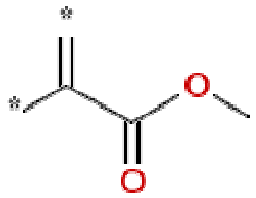
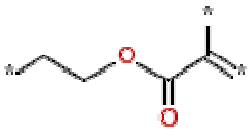
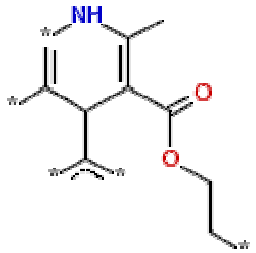
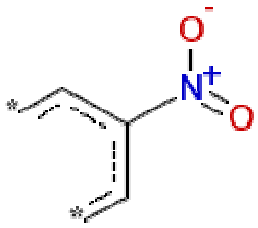
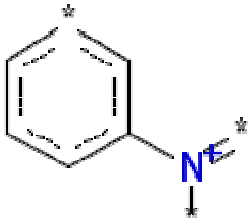
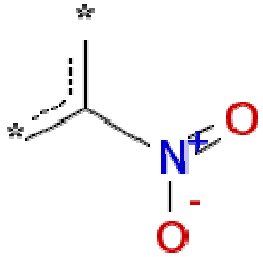
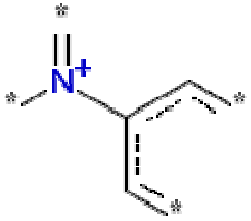
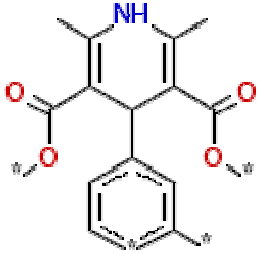
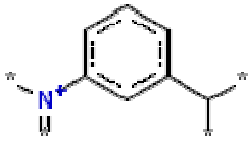
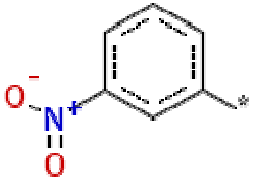
This table shows, for each category, statistics derived from the cross-validated predictions of the model built for that category as applied to members of that category and non-members of that category. For each group, the number of members/nonmembers (N) is given; the mean prediction for each subset (Mean); and the estimate standard deviation of the predictions for each subset (StdDev).

Category N	Category Mean (±StdDev)	Noncategory N	Noncategory Mean (±StdDev)
17	6.02 (±8.34)	21	-4.70 (±3.95)

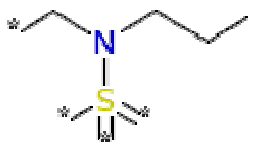
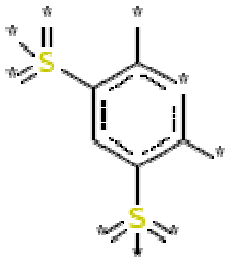
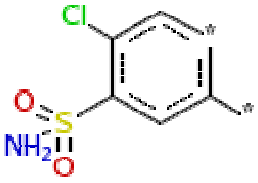
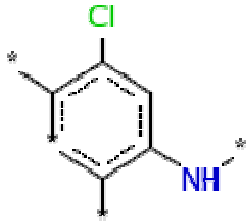
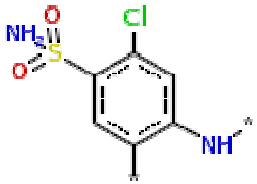
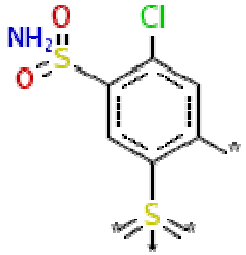
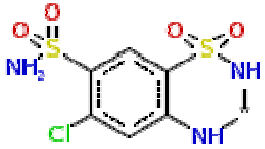
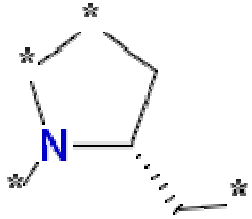
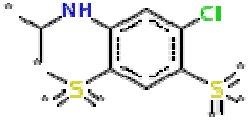
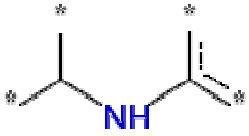
Supplemental Table S7. ASBT binding features from Bayesian analysis.

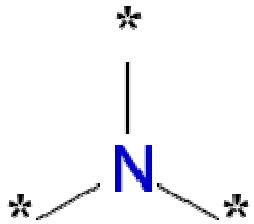
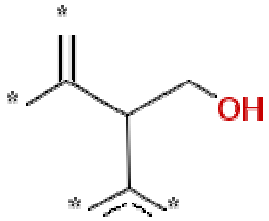
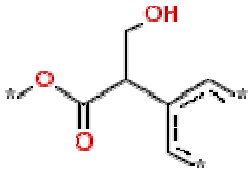
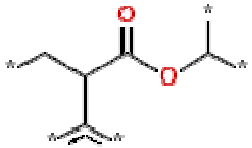
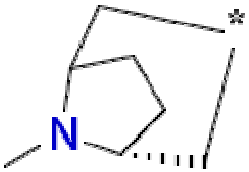
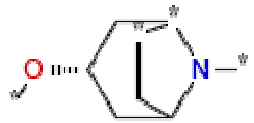
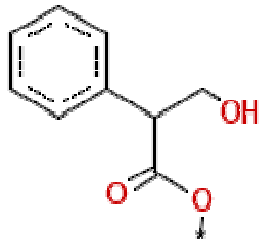
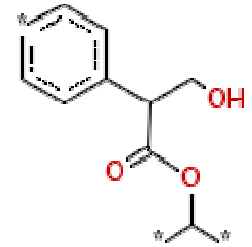
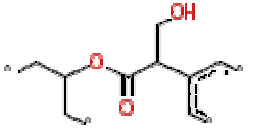
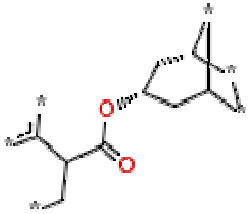
A. Favorable features for inhibition

				
G1: 436975416 7 out of 7 good Bayesian Score: 0.609	G2: 742790379 7 out of 7 good Bayesian Score: 0.609	G3: 1316670154 7 out of 7 good Bayesian Score: 0.609	G4: -2045589248 7 out of 7 good Bayesian Score: 0.609	G5: 153483420 7 out of 7 good Bayesian Score: 0.609
				
G6: -1481289364 7 out of 7 good Bayesian Score: 0.609	G7: 325836504 7 out of 7 good Bayesian Score: 0.609	G8: 1766339474 7 out of 7 good Bayesian Score: 0.609	G9: -2128664250 6 out of 6 good Bayesian Score: 0.592	G10: -174436305 6 out of 6 good Bayesian Score: 0.592

 G11: 339830961 6 out of 6 good Bayesian Score: 0.592	 G12: -432846198 5 out of 5 good Bayesian Score: 0.570	 G13: -2127475941 5 out of 5 good Bayesian Score: 0.570	 G14: 1579401580 4 out of 4 good Bayesian Score: 0.539	 G15: -101285563 4 out of 4 good Bayesian Score: 0.539
 G16: -1338588315 4 out of 4 good Bayesian Score: 0.539	 G17: -828984032 4 out of 4 good Bayesian Score: 0.539	 G18: -797095437 3 out of 3 good Bayesian Score: 0.496	 G19: -405875953 3 out of 3 good Bayesian Score: 0.496	 G20: -292898213 3 out of 3 good Bayesian Score: 0.496

B. Detrimental features for inhibition.

 B1: 309602933 0 out of 4 good Bayesian Score: -1.070	 B2: -451251206 1 out of 9 good Bayesian Score: -0.976	 B3: 48712700 0 out of 3 good Bayesian Score: -0.891	 B4: -770645118 0 out of 3 good Bayesian Score: -0.891	 B5: 1198988172 0 out of 3 good Bayesian Score: -0.891
 B6: -1800409220 0 out of 3 good Bayesian Score: -0.891	 B7: 834375811 0 out of 3 good Bayesian Score: -0.891	 B8: -1946918893 0 out of 3 good Bayesian Score: -0.891	 B9: 851915 0 out of 3 good Bayesian Score: -0.891	 B10: 1294255210 1 out of 7 good Bayesian Score: -0.778

				
B11: 9 2 out of 11 good Bayesian Score: -0.737	B12: -1094445514 0 out of 2 good Bayesian Score: -0.672	B13: 1424304659 0 out of 2 good Bayesian Score: -0.672	B14: -214983127 0 out of 2 good Bayesian Score: -0.672	B15: 1380384081 0 out of 2 good Bayesian Score: -0.672
				
B16: -987903557 0 out of 2 good Bayesian Score: -0.672	B17: -1824082254 0 out of 2 good Bayesian Score: -0.672	B18: 1154116349 0 out of 2 good Bayesian Score: -0.672	B19: 365588023 0 out of 2 good Bayesian Score: -0.672	B20: 1649104107 0 out of 2 good Bayesian Score: -0.672

Supplemental Table S8. Validation of analysis of the quantitative model without the excluded volume.

	True positives	False negatives	False positives	True negatives
Training set (n=38)	12 (31.6%)	5 (13.2%)	0 (0%)	21 (55.3%)
Testing set (n=30)	6 (20.0%)	5 (13.3%)	5 (16.7%)	14 (50.0%)
Literature testing set (n=19)	7 (36.8%)	8 (42.1%)	0 (0%)	4 (21.1%)